# Knowledge and Growth: Evidence From Early Modern Europe

Francisco Queiró, Nova School of Business and Economics *

2018

**Abstract**

I investigate the relationship between knowledge diffusion and city growth using a new dataset of 5.5 million books published in Europe from 1450 to 1800. The dataset consists of individual book data drawn from over 72,000 library catalogs around the world, including most major national and research libraries. Exploiting within-city variation, I find that book production is a strong predictor of subsequent population growth. Splitting books by subject, I find that the results are robust for books on technology, finance, medicine and history, with technology and finance having the largest coefficients. In addition, although science books as a whole are insignificant, books on chemistry and geology also increase growth, which is consistent with the important roles of chemistry and coal mining during the Industrial Revolution. Books on other topics, such as religion or literature, are not associated with growth.

*I am very grateful to my advisors Andrei Shleifer, Lawrence Katz, Josh Lerner and Alberto Alesina. I also thank Ruchir Agarwal, Claudia Goldin, Arash Nekoei, Nathan Nunn and Luigi Zingales for comments and guidance.

# I Introduction

Human capital may affect economic growth through different channels. One influential view, going back to the work of Nelson and Phelps (1966) and Schultz (1975), is that more educated people are more adaptable, and in particular that they absorb new knowledge and adopt new technologies faster. Given the difficulty of measuring knowledge diffusion, evidence on this hypothesis has been mostly indirect.[1]

One idea for a broad measure of knowledge diffusion, introduced by Baten and van Zanden (2008), is book production. Yet while books are a key tool in the transmission of economically useful knowledge, they also serve other purposes, such as the diffusion of religious and cultural norms, political debate, artistic expression or entertainment. The aggregate book counts used by Baten and van Zanden (2008) are therefore open to different interpretations. They could be a better proxy for the level of literacy or consumption than for the type of knowledge economists would describe as technology.

This paper investigates the relationship between knowledge diffusion and growth at the city level using a new dataset of 5.5 million books published in ten European countries[2] from 1450 to 1800. The dataset consists of individual book data as recorded by WorldCat, a network of over 72,000 library catalogs around the world. The WorldCat network includes the national libraries of the ten countries in the sample, among other major national and research libraries, which implies that its combined catalogs plausibly contain all surviving book titles published in these countries. Crucially, the data contain information on the book's subject for a sample of 12 percent of book titles, enabling me to distinguish economically useful knowledge – e.g. books on technology – from literacy or consumption – e.g. religious books and literature.

The dataset covers the so-called hand-press period, between Gutenberg's invention of the printing press and the introduction of industrial printing presses in the early 19th century, and the historical context is important for several reasons. First, unlike in today's digital

[1] Barro (1991) and Benhabib and Spiegel (1994) find that initial schooling predicts economic growth, as implied by the Nelson and Phelps (1966) model. Ciccone and Papaioannou (2009) find that initial schooling predicts faster growth in schooling-intensive industries when technological progress is skill-biased.

[2] The dataset covers western European countries whose national library is part of the WorldCat network: Denmark, England, France, Germany, Ireland, Netherlands, Scotland, Spain, Sweden and Switzerland. Countries are defined by their modern borders.

age, book production was arguably the main medium for knowledge transmission in this period. Second, local book production was a reasonable proxy for the local demand for knowledge. Dittmar (2011) presents evidence that transport costs for books were high in this period, and that it was more common for books to spread through reprints than inter-city trade. Schmitt et al. (1988) adds that lack of copyright protection kept print runs small and also encouraged local reprints as a diffusion mechanism.[3] Third, the data include separate records for different "manifestations" of a book, which include new editions and reprints. This implies that differences in diffusion across books are well captured by the data.[4] Finally, the use of an extended time period allows me to account for long lags in the effect of knowledge diffusion on output while still observing significant variation over time, which allows me to identify this effect from within-city variation alone.

I start by presenting basic trends in book production over time, both across countries and subjects. Switzerland and Germany were the early leaders in printing, both playing an important role in the diffusion of the Reformation in the early 16th century. Germany experienced a sharp fall in production during the Thirty Years' war (1618-1648), and England and the Netherlands rose to the top in the 17th century. France was behind the leading countries throughout most of the period, and caught up at the onset of the French revolution, while Spain had low levels of production throughout the entire period. Scotland, Denmark and Sweden were somewhere between the leading countries and France, and Ireland had virtually no production until the 17th century but grew steadily after that. Across subjects, early printing was dominated by religion and literature. A clear spike in religious books is visible after the publication of Luther's Ninety-Five Theses in 1517. Over time, some subjects – history, science, medicine, law – saw their share increase steadily, while other subjects only took off in the 18th century, such as technology, business and finance, agriculture and social science.

I then turn to examining the relationship between book production and city growth. I

[3] An an example, Schmitt et al. (1988) notes that "from 1509 to 1520, Erasmus' *Moriae encomium* appeared in at least thirty-five editions in nine different cities, printed by fourteen or more publishers. The places of publication were Antwerp, Basle (two publishers), Cologne, Florence, Mainz, Paris (four publishers), Sélestat, Strasburg and Venice (two publishers). Reprinting was almost routine at a time when the concept of literary property or binding commercial restrictions hardly existed".

[4] This point is also made by Schmitt et al. (1988), who writes that "the number of reprints, if any, presents a truer measure of a book's diffusion [than initial print runs]".

use population growth as a measure of economic growth at the city level, following standard practice in both historical and modern contexts (De Long and Shleifer, 1993; Glaeser, Scheinkman and Shleifer, 1995).[5] Exploiting differential within-city variation, and accounting for shocks at the country level, I find that book production has a strong effect on subsequent population growth. The results are robust to a variety of specifications, including alternative assumptions about the rate of incorporation of knowledge into output and the inclusion of city-specific time trends. The coefficient is also stable across time periods and robust to splitting the sample by country, with the exception of England where publishing was exceptionally concentrated in London.

Next, I distinguish between different interpretations of this effect by splitting books by subject. Books on technology, finance, medicine and history are robustly related with growth, with technology and finance having the strongest effect. The first three subjects clearly capture the production and diffusion of economically useful knowledge. History includes books about current events at the time they were written, and one interpretation is that it captures the existence of a literate elite involved in or at least informed about public life. In addition, I show that although science books as a whole are insignificant, books in chemistry and geology in particular also increase growth in this period, with a coefficient similar in magnitude to that of technology and finance. This is consistent with the importance of the 18th-century Chemical Revolution (Clow and Clow, 1952) and coal mining as drivers of the Industrial Revolution. Books on religion or literature, which comprise the majority of the sample, are not robustly associated with growth. Overall, these findings suggest that the effect of book production reflects the diffusion of knowledge rather than literacy or consumption.

The findings in this paper contribute to an emerging empirical literature that emphasizes the importance of upper-tail human capital, and educated entrepreneurs in particular, as a driver of economic growth through faster technology adoption, in line with the views of Nelson and Phelps (1966) and Schultz (1975). Mokyr (2005*b*) distinguishes the roles of average and upper-tail human capital in the context of the industrial revolution, and argues

[5]The underlying logic is that productivity shocks trigger migration towards more productive cities to the point that potential migrants remain indifferent across locations, with rising house prices limiting the extent of migration (Roback, 1982; Glaeser and Gottlieb, 2009)

that the latter – scientifically knowledgeable entrepreneurs and engineers – was key in the diffusion and application of useful knowledge in production. Squicciarini and Voigtländer (2014) provide evidence in favor of this view by showing that the density of encyclopedia subscriptions in French cities, a proxy for the presence of upper-tail human capital, increased city growth from 1750 to 1850, although they find no effect on earlier growth. In a modern context, Gennaioli et al. (2013) show that the education of managers can account for substantial variation in productivity across firms and regions.

The paper also contributes specifically to the literature on the historical role of knowledge diffusion and human capital before and during the Industrial Revolution. Dittmar (2011) shows that early adoption of the printing press had a sizable effect on city growth in the 16th century. In independently developed work, Dittmar (2015) identifies book subjects in a different sample of books published between 1450 and 1600, and finds that business books in particular were correlated with growth in the cross-section of cities. The period covered by Dittmar's analysis, however, excludes the Scientific Revolution and Enlightenment of the 17th and 18th centuries, both widely regarded as key periods in the advancement of knowledge. Cantoni and Yuchtman (2014) find that the creation of universities increased growth, as measured by the establishment of markets, in medieval Germany.

The predominant view among historians is that scientific knowledge (Mathias, 1969; Hall, 1974) and human capital (Mitch, 1993; Allen, 2003; Galor, 2005) did not become engines of growth before the second phase of industrialization, in the second half of the 19th century. Another view (Musson and Robinson, 1969; Mokyr, 2005*a*) argues that, while scientific breakthroughs might have played a limited role in this period, the scientific mindset that spread among elites in the 18th century set off an accumulation of useful knowledge, consisting of "catalogs of facts, based on experience and experiment rather than on understanding or careful analysis and testing" (Mokyr, 2005*a*), which in turn drove industrialization.

This paper offers support for this second view. First, it confirms that there was a sharp increase in books on science and technology in the 18th century. Second, it shows that while some scientific fields – notably physics and biology – did not increase growth, in line with the predominant view, books on technology, chemistry and geology did have a strong effect on growth in this period. Third, it finds that knowledge increased growth long before the

Industrial Revolution, in line with Dittmar (2011). And fourth, it provides evidence on the effect of knowledge and upper-tail human capital on city growth by exploiting within-city rather than cross-sectional variation as in Dittmar (2011), Dittmar (2015) or Squicciarini and Voigtländer (2014).

Lastly, the paper also relates to the literature on the history of book production. Relative to Baten and van Zanden (2008) and Buringh and Van Zanden (2009), who report a total of just under 630 thousand books published in all of Europe in this period, this paper introduces significantly more comprehensive data on book production, covering 5.5 million books published in the ten countries in the sample alone and using a single source of data.[6]

The rest of the paper is organized as follows. Section II describes the data. Section III presents trends in book production. Section IV outlines the empirical methodology and reports the main findings. Section V concludes.

# II Data

## *II.A Book Records*

The paper uses a new data set constructed from book records drawn from WorldCat, a collection of over 72,000 library catalogs in 170 countries, including most major libraries around the world.[7] WorldCat is produced and maintained by the Online Computer Library Center cooperative, and it is the world's largest bibliographic database, with over 330 million records of books, periodicals, visual materials and sound recordings, among other document types.

Each record corresponds to a "manifestation" of a work. A manifestation could be an original title, a new edition or a reprint. WorldCat assigns a unique identifier to each record, and implements matching algorithms to eliminate duplicates. Among document types, books are defined as "books, pamphlets, technical reports, typescripts, theses, dissertations, manuscripts and other written works".

[6] Baten and van Zanden (2008) and Buringh and Van Zanden (2009) use multiple sources and adjust book totals to correct for perceived differences in the comprehensiveness of each source.

[7] WorldCat covers 17 of the 19 largest libraries in the world, including all top 10, as listed in Wikipedia. (https://en.wikipedia.org/wiki/List_of_largest_libraries)

Book records contain information on the book's title, author, place and year of publication and language among other details. Importantly for the purposes of this paper, some book records contain information on the book's subject, which will be described in greater detail below. For illustration, figure 1 displays the record for the first major book printed in Europe, Gutenberg's Bible. The record shows that this book was printed in Mainz in either 1454 or 1455. Whenever a record indicates a range rather than a specific year of publication, as in this case, I assign an equal fraction of the book to each year in the range when computing annual book totals.

### *II.B City Population*

City population comes from Bairoch, Batou and Pierre (1988), henceforth the Bairoch data set, following standard practice in studies of city growth in early modern Europe. This data set covers 2,204 European cities that had a population of at least five thousand at any point between 800 and 1800, and it reports city population, when available, every 100 years up to 1700 and every 50 years after that until 1850.

### *II.C Sample Definition*

I focus the analysis on books published between 1450 and 1800 in Western European countries whose national library is part of WorldCat. This implies that the combined WorldCat catalogs plausibly contain all surviving book titles published in these countries. The ten countries that meet this criteria are Denmark, England, France, Germany, Ireland, Netherlands, Scotland, Spain, Sweden and Switzerland.[8] This leads to a set of just over six million book records. From each record, I collect information on the book's country, place and year of publication, and on its subject, when available.

[8]One important omission is Italy, an important center of early printing (Dittmar, 2011) whose national libraries of Rome and Florence are absent from WorldCat at present.

## II.D Matching Books and Cities

I then match these records with the cities in the Bairoch data set. There are two main challenges in the matching process. First, many cities appear under multiple names in book records, including variations in spelling and names in different languages. Latin city names, for example, are common. The Bairoch data includes a set of alternative city names which I also use in the matching process, but these only mitigate the problem slightly. I overcome this issue by using variant city names from the Consortium of European Research Libraries (CERL) Thesaurus, a dataset that contains variations in the names of print locations up to the mid-nineteenth century. To illustrate the severity of the problem, the city of Berlin has 58 alternative names in the CERL Thesaurus data, some of which are recognizable, like Berlinium, while others are less so, like Colonia Marchica. The second challenge is that city names are often misspelled, or appear within longer strings of text in the record's place of publication field. In addition, a few records indicate multiple cities, which I take to mean that the book was published in all cities mentioned. To address these issues, I take every sequence of up to four words in the place of publication field and match it with city names using a fuzzy string matching procedure based on the Ratcliff-Obershelp algorithm[9]. See appendix for details. If there are multiple city matches for a given book record, then I assign that book to each of the matched cities.

The matching procedure yields at least one city match for 5.5 million records, or 91 percent of all books published in the 10 countries in the sample, and these records constitute the analysis sample for this paper. The unmatched nine percent fall almost exclusively under two categories: either the place of publication is missing, or the city and country reported do not match, e.g. the country of publication field reports "France" while the place of publication field reports "Amsterdam".[10] Overall, the outcome of the matching process

[9]This algorithm computes the similarity of two strings as twice the number of matching characters divided by the total number of characters in each string. Matching characters are those in the longest common subsequence of characters plus, recursively, matching characters in the unmatched region on either side of the longest common subsequence. I implement this algorithm using a Python package named FuzzyWuzzy, which is a wrapper for the difflib module in Python's standard library.

[10]Regarding the second case, some controversial books in this period were published under fictitious publisher names from a different city, often abroad, in order to protect the real publishers. A famous case is that of the publisher Pierre Marteau from Cologne. It was invented by exiled French printers in the Netherlands in the 17th century, and used by publishers in France and the Netherlands, and later in Germany, mostly in order to publish political satire (de Brouillant, 1888). In the 18th century a supposed

confirms that book printing was highly concentrated in cities, as argued by Dittmar (2011).

### *II.E Book Subjects*

A subset of book records include classification codes from standard library classification systems. The classification systems present in the data are the Library of Congress Classification (LCC), the Dewey Decimal Classification (DDC), the Universal Decimal Classification (UDC), the National Library of Medicine classification (NLM) and the National Agricultural Library classification (NAL). For example, the book record in figure 1 reports the LCC code BS75, which corresponds to class B "Philosophy, Psychology, Religion" and subclass BSThe Bible". Availability of these codes is presumably related to whether the library contributing the record uses one of these systems or not. Academic libraries in the U.S. and U.K. tend to use the LCC system, while public and school libraries tend to use DDC. In continental Europe, libraries tend to use DDC or UDC, but the use of standard classification systems is significantly less widespread than in the U.S. and U.K.

Overall, 12 percent of book records in the sample include at least one code from either the LCC, DDC or UDC systems. The NLM and NAL would add an additional one percent, but I exclude these subject-specific systems to avoid biasing the subject composition in the sample.[11] The most commonly reported code is the LCC, covering 10 percent of records. The DDC is next with 2 percent, and the UDC contributes less than 1 percent. When more than one code is available for a given record I take the LCC if available, and the DDC otherwise.

I use these codes to define 14 subjects: religion, literature, arts, history, language, philosophy, general, science, medicine, technology, agriculture, social science, business and finance, and law. These subjects correspond broadly to LCC, DDC and UDC first or second level divisions. Table 1 provides the correspondence between subjects and LCC, DDC and UDC codes. Subject labels are self-explanatory with a couple of exceptions. The most important one is history. Many books under this label were accounts of current events when they were

line of family members kept the label flourishing, with books published by Marteau's widow and heirs. In light of this, it is possible that for some books the correct country of publication is known and reported, along with the fictitious publisher and city. A WorldCat search for books published by Pierre Marteau in the Netherlands and France yields several examples of such cases.

[11] Including them does not change any of the results in the paper.

published. As will be seen below, there were spikes in publications classified as history during major historical events like civil wars and revolutions. It is therefore more appropriate to think of this label as a combination of history and current events. The general label includes reference works, such as dictionaries, encyclopedias, indices, almanacs, bibliographies, and collections of periodicals.

Subject coverage is not even across countries. In particular, because libraries in the U.S. and U.K. are more likely to use standard classification systems, books published in England, Scotland and Ireland are more likely to have subject information than others. For example, 21 percent of books published in England have subject information, while in France and Germany that number falls to 11 and 9 percent respectively. Year of publication, on the other hand, is only weakly related to subject availability,[12] but books published during particular historical episodes, such as the English civil war, appear to be more likely to have subject information. In order to account for these biases in subject coverage, I adjust all subject-specific book counts by multiplying the raw count by the ratio of total books to books with subject information at the country-year level.

# III Trends in Book Production

In order to summarize the data, I start by presenting trends in book production, first across countries and then across subjects.

## *III.A Cross-Country Trends*

Figure 2 displays annual per capita book production by country for the entire sample period.[13,14] For visibility, I split the ten countries in the sample into two graphs. The top

[12] In some countries the relationship is positive and in others it is negative, but in all cases its magnitude is small.

[13] The numerator includes all books produced in each country, including the nine percent of books I am unable to match with a city. It is highly likely that these books were also published in cities in the Bairoch data, see section II for a discussion of unmatched books. Including only matched books makes little difference for the patterns observed in the graphs.

[14] The denominator includes only urban population, as measured in the Bairoch data, rather than the population of the entire country. Since printing was almost exclusively urban, this seems to be the appropriate choice conceptually. Using total country population only makes a substantial difference for the Netherlands, which had a much higher urbanization rate than other countries in the sample, and a smaller difference for

graph includes the five largest countries by population, and the bottom graph includes the remaining five countries. I use a log scale for books, both to reduce the visual impact of outliers, such as England in the early 1640s, and to emphasize differences across countries at different points in time.

Switzerland had an early leading role as a printing center for the Reformation, which persisted until the early 17th century. Calvin and Zwingli were both based in Switzerland and made heavy use of the printing press to disseminate their ideas. A clear spike in book production is visible in 1517, the year Luther's Ninety-Five Theses were published. Germany, where printing technology was invented, was also particularly active in the early period. The same spike in production is visible in 1517, although there it was short-lived, but Germany was among the leaders until the onset of the devastating Thirty Years' war (1618-1648). Book production experienced a sharp drop in this period, and only fully recovered in the 18th century. As Germany collapsed, England rose to the top with a printing explosion during the civil war in the early 1640s. Book production remained high and volatile throughout the revolutionary period, and then declined somewhat in the 18th century, at which point other countries caught up with England.[15] The Netherlands experienced a significant increase after independence from Spain in the late 16th century, and remained just below England until the very end of the sample, with book production dropping in the 1790s. France was initially among the leaders but fell behind in the 1560s. The brief spike around 1650 corresponds to the Fronde, a series of civil wars between 1648 and 1653. It remained behind until the revolution in 1789, which led to a burst of publications. Spain, unlike the rest of Europe, had very low levels of book production throughout the period.

Among the remaining countries, in the bottom graph of figure 2, Scotland, Denmark and Sweden were somewhere between France and the leading countries for most of the sample period, and caught up with the leaders in the 18th century. Scotland had a period of intense publishing in the decades around 1700, which coincides with the start of the Scottish Enlightenment. Ireland had virtually no book production until the 17th century; in the 18th

Germany, which had a lower urbanization rate. Annual population numbers for these graphs are obtained by interpolating the Bairoch data and assuming constant population outside the interval of interpolation.

[15]It has been argued that England had lower levels of literacy than Sweden or Germany at the onset of the Industrial Revolution (Mitch, 1993), as measured by the ability to sign documents. This was not true in the case of book production, however, as Baten and van Zanden (2008) and this paper both show.

century its production grew substantially but it remained behind the leading countries.[16]

### III.B Subject Trends

A key advantage of these data is the ability to classify a sample of books by subject. Figure 3 offers a visual overview of subject trends in this sample by showing the share of books by subject over time, where I aggregate the 14 subjects defined in section II into six broader categories.

Early printing was dominated by religion and literature. Religious books comprised up to 80 percent of books in the first decade of printing. Between the late 15th century and the first half of the 17th century, they represented about 35 percent of books, with the exception of the 1520s, at the onset of the Reformation, when they rose to 80 percent again. After this the share of religious books dropped markedly, falling to below 10 percent by 1800. Literature, arts and language also thrived in the 15th and 16th centuries, driven by the publication of Greek and Roman classics. Roughly 40 percent of books fell into this category between the 1460s and 1510. After the 1520s, this fraction remained relatively constant at around 30 percent throughout the sample period.

Other subjects gained importance over time. History, which includes books on current events as explained in section II, took off in the mid-16th century and represented about 20 percent of publications from then onwards, with large spikes during major historical episodes such as the English and French civil wars of the mid-17th century and the French revolution in 1789. The share of books on law, business and social science grew steadily throughout the sample period, from about five to 20 percent. Science and technical books, which include books on science, medicine, technology and agriculture, are closely associated

[16] Baten and van Zanden (2008) and Buringh and Van Zanden (2009) also present data on the evolution of book production across countries in the same period. They report only 50-year aggregates rather than annual data, so a detailed comparison cannot be made. Still, the findings presented here appear to be broadly in line with theirs. The major exception is that they report much higher book production per capita in the Netherlands than in all other countries from the 17th century onwards. This difference is at least partially explained by the fact that they use total population in the denominator, whereas I use urban population only, and the Netherlands had a higher urbanization rate than the rest of Europe in this period. They also find that production dropped sharply in Switzerland in the 17th century, to the level of Spain, and that Sweden experienced a large increase in production in the second half of the 18th century, making it the second highest book producer in per capita terms. In the WorldCat data Switzerland remained close to the leaders; Sweden grew steadily over time and was among the leaders at the end of the 18th century, but did not stand out.

with economically useful knowledge. These books represented a small but growing share of publications throughout the sample period, going from about five percent in the 15th century to about 12 percent in the late 18th century. Finally, the "other" category includes philosophy and general books, and remained below five percent of books throughout the period.

Figures 4 and 5 offer a more detailed view by showing per capita book production for each of the 14 subjects. Importantly for the analysis below, they offer a breakdown of science and technical subjects from figure 3. Within this category, books on science and medicine grew from the 16th century onwards, while books on technology and agriculture only took off in the second half of the 18th century, at the onset of the Industrial Revolution. Science publications also accelerated markedly in the 18th century, as did those in social science, business and finance, and arts, while religious books declined. This is consistent with the Age of Enlightenment associated with this period.

How correlated was the publication of books across subjects? Table 2 presents the correlation matrix for log books per capita in each of the 14 subjects at the city-year level. Publications were positively but not strongly correlated across subjects, with most pairwise correlations falling in the 0.15-0.45 range.

# IV The Effect of Book Production on City Growth

## *IV.A Methodology*

Population is a widely used measure of economic development at the city level, both historically and in modern contexts (De Long and Shleifer, 1993; Acemoglu, Johnson and Robinson, 2005; Glaeser, Scheinkman and Shleifer, 1995). The underlying logic is that differences in productivity growth across cities trigger migratory flows such that the utility of potential migrants remains equalized across cities, with rising house prices limiting migration to more productive cities (Roback, 1982; Glaeser and Gottlieb, 2009).[17] Following this approach, let

[17] Blanchard and Katz (1992) show that this mechanism is also active across U.S. states

city growth be given by

$$\Delta \ln L_{it} = \Delta \ln A_{it} + \alpha_i + \eta_{jt} + \varepsilon_{it} \quad (1)$$

where $\Delta$ denotes the change from $t$ to $t+1$, $L_{it}$ is population, $A_{it}$ is the level of technology, $\alpha_i$ captures fixed city-level drivers of growth, such as geographic features, persistent institutions or culture, $\eta_{jt}$ are country-level shocks, such as revolutions or wars, and $\varepsilon_{it}$ are city-level shocks.

Technological change reflects the incorporation of new knowledge, which will be measured by per capita book production $\frac{B}{L}$, into the existing technological stock. New knowledge is presumably not incorporated instantly into production processes, but instead diffuses over time. I assume that new knowledge is incorporated into technology at the rate $\gamma$ and that the growth rate of technology can be written as

$$\Delta \ln A_{it} = F\left[\gamma \sum_{s=0}^{t} \frac{B_{is}}{L_{is}} (1-\gamma)^{(t-s)}\right] \quad (2)$$

The sum corresponds to the stock of knowledge that has not been incorporated into technology at time $t$, and $F$ is an increasing function capturing the productivity of new knowledge. This equation parallels the technological change equation in knowledge-based models of endogenous growth (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992), with new knowledge explicitly measured by books per capita instead of by the amount of human capital allocated to research.

Assuming a log form[18] for $F$ and plugging (2) into (1), city growth can be expressed as a function of the stock of unincorporated knowledge at $t$ and unobserved factors $\alpha_i$, $\eta_{jt}$ and $\varepsilon_{it}$:

$$\Delta \ln L_{it} = \beta \ln \left[\sum_{s=0}^{t} \frac{B_{is}}{L_{is}} (1-\gamma)^{(t-s)}\right] + \alpha_i' + \eta_{jt} + \varepsilon_{it} \quad (3)$$

where $\alpha_i' \equiv \alpha_i + \beta \ln \gamma$. Equation (3) will be the main estimating equation in the analysis below.

[18] The distribution of per capita book production is considerably right-skewed. Figure 6 suggests that a linear relationship between log city growth and log unincorporated knowledge is a good approximation. I use the log of one plus unincorporated knowledge in order to include city-year observations without any book production in the analysis.

In the Bairoch data, population after 1450 is available in 100-year intervals between 1500 and 1700 and in 50-year intervals between 1700 and 1850. I therefore measure $\Delta \ln L_{it}$ as $L_{it+100} - \ln L_{it}$ in the earlier period and $2 \times (\ln L_{it+50} - \ln L_{it})$ in the later period, so that all growth rates are expressed in the same units.

The measurement of unincorporated knowledge $\sum_{s=0}^{t} \frac{B_{is}}{L_{is}}(1-\gamma)^{(t-s)}$ requires assuming a value for $\gamma$ and a value for the initial stock in 1450. For my main specification I choose $\gamma = 0.1$, which implies that over 90 percent of knowledge is incorporated into technology within 25 years (i.e. $1 - 0.9^{25} = 0.928$), but I also show that the results are not very sensitive to other choices within a reasonable range for $\gamma$. The value of the initial stock in 1450 is unimportant given reasonable values of $\gamma$, and I set it to 0.[19] In addition, the measurement of $\frac{B}{L}$ requires annual population estimates, which I obtain by linear interpolation of log population using the Bairoch data.[20]

## *IV.B Results for All Books*

I start by presenting results from estimating (3) for all books, without distinguishing between subjects. Throughout the rest of this section and in the corresponding graphs and tables I refer to unincorporated knowledge, as defined in equation (2), as books or book production for simplicity. In addition to the city and country-by-year fixed effects in (3), the set of controls for all regressions includes a quartic in log population at $t$, and standard errors are clustered at the city level.

Figure 6 presents the relationship between city growth and log book production in my baseline specification, which sets $\gamma = 0.1$. The figure shows a binned scatter plot of city growth and log book production, both residualized on the set of controls, along with the corresponding regression line, coefficient and standard error.[21] The coefficient on log book

[19] For $\gamma = 0.1$, less than one percent of the knowledge created before 1450 was unincorporated into technology by 1500, the time of the first growth observation in the data. In any case, book production before printing was extremely low. According to Buringh and Van Zanden (2009) there were more books produced in Europe between 1450 and 1500 than in the preceding one thousand years.

[20] Outside the interval of known population points I assume constant population equal to the closest known point. For example, if the Bairoch data reports population between 1600 and 1750 only, I interpolate population between 1600 and 1700 and assign the 1600 population to years before 1600 (assigning population after 1700 is not necessary, since the last city growth observation is the one between 1700 and 1750).

[21] The plot shows mean residual city growth and log book production for 19 equal-sized bins for all cities with positive book production at any point during the sample period plus one bin including all cities without

production equals 8.43 log points, and is highly significant. In addition, the plot shows that the relationship is approximately linear, suggesting the log form for book production is an appropriate choice.

Table 3 presents the corresponding regression in column one, along with variations on the baseline specification in the remaining columns. Columns two and three show that the coefficient is not very sensitive to different values of $\gamma$. In column two I use $\gamma = 0.2$, which implies that over 90 percent of knowledge is incorporated in around 12 years instead of 25, and the coefficient rises slightly to 9.65 log points. In column two I use $\gamma = 0.05$, in which case it takes around 50 years for the same 90 percent of knowledge to be incorporated into technology, and the coefficient falls slightly to 7.14 log points. Column four adds linear city-specific time trends, to further account for unobserved drivers of growth. This is a particularly demanding robustness test since there are only five observations per city at most, with an average of three. The linear trends therefore absorb a considerable amount of variation in growth and book production and amplify the effect of measurement error. In this specification the coefficient falls to 5.56 log points and remains significant at the 10 percent level. Column five restricts the sample to cities with positive book production, and the coefficient is essentially unchanged at 8.64 log points. Column six weights observations by average city population across the sample period, which increases the coefficient to 11.12 log points. Finally, column seven shows that the results are robust to using levels instead of logs in book production.

Next, I examine how this relationship varies across time and geographies. The role of knowledge and human capital as drivers of economic growth before industrialization is disputed. While Dittmar (2011) and Cantoni and Yuchtman (2014) show that adoption of the printing press and the creation of universities, respectively, increased growth in late medieval and early modern Europe, Squicciarini and Voigtländer (2014) argue that the presence of knowledge elites did not increase growth before 1750. Column one in table 4 interacts log book production with time indicators and shows that book production increased growth both before and during industrialization. The coefficient on books is stable and highly significant

book production. The regression line, coefficient and standard error are estimated on the underlying data, not the binned averages. All other binned scatter plots in the paper follow the same procedure.

across periods, ranging from 8.95 to 12.44 log points, with the exception of the coefficient for 1750, which is small and insignificant.[22] These results therefore do not reject the idea that knowledge and human capital were unimportant right at the onset of industrialization, in the second half of the 18th century, but suggest that if anything this period was exceptional in that regard, rather than representative of the pre-industrial age.

The remaining columns in table 4 split the sample by geography. Columns two to six present regressions for each of the five largest countries by population – France, Germany, England, Netherlands and Spain. Column seven groups the smaller countries – Denmark, Ireland, Scotland, Sweden and Switzerland – into one regression.[23] The coefficient on book production is significant in all cases except England, and ranges from 7.36 in the smaller countries to 15.64 log points in France. The fact that England is not significant is unsurprising, because publishing in England was uniquely concentrated. Printing was legally restricted to London, with the exception of Oxford and Cambridge Universities, from the 16th to the early 18th century. A Royal Charter in 1557 gave the Stationers' Company, a guild of printers in London, a monopoly on printing and the legal power to enforce it. The Statute of Ann in 1710 instituted copyright protection for authors and ended the Stationers' monopoly, but London's dominance persisted. The book production data show that these restrictions were highly effective: 93 percent of books published in England during the sample period were published in London, followed by Oxford and Cambridge with 2 and 1 percent respectively. This suggests that local book production was a poor proxy for local consumption in English cities, and that the English data probably have a low signal to noise ratio. Other capitals also had a prominent role as publishing centers, but not as dominant.[24] All results are unchanged when capital cities in all ten countries are excluded.

To get a sense of the magnitude of these findings, a city in the $75^{th}$ percentile of book production among all city-years with positive book production grew approximately 17.4 log

[22]The same pattern holds in cross sectional regressions of city growth and book production by time period, although the coefficient in the regression for 1750 is slightly larger and closer to being significant.

[23]There are too few observations in the smaller countries to estimate the effect by country.

[24]The highest level of concentration after England was in France, where 74 percent of books were published in Paris. In Spain, 47 percent of books were published in Madrid and in the Netherlands 45 percent were published in Amsterdam. Among the five largest countries, Germany had the least concentrated book production, probably because it was not politically integrated in this period. The top producers in Germany were the two cities that held major book fairs, Leipzig and Frankfurt, with 20 and 9 percent respectively.

points faster than a city without book production. This compares with an average growth of 52.8 log points and a standard deviation of 87.7 log points.

## *IV.C Results by Subject*

The findings presented so far could be interpreted in different ways. Book production could measure knowledge diffusion, but it could also measure the level of literacy or consumption. In fact, section III showed that the majority of books published in this period were either on religion, literature, language or arts. This section distinguishes between these interpretations by estimating the effect of book production on growth by subject, using the 12 percent sample of book records with subject information. All estimates assume $\gamma = 0.1$, although again the results are not very sensitive to this choice.

Column one in table 5 reports the coefficients from estimating equation (3) separately for each of the 14 subjects defined in section II. Given that book production is positively correlated across subjects, as shown in table 2, this specification minimizes bias from measurement error but suffers from omitted variable bias, and is therefore loaded in favor of each subject. Seven subjects have positive and significant coefficients in column one. Books on business and finance and books on technology have the highest coefficients: 21.5 and 17.2 log points respectively. Next come books on medicine and history, with 10.0 and 9.3 log points, followed by arts with 8.6, social science with 7.2 and finally religion with 4.2. The coefficients on technology and history are significant at the 1 percent level, while the coefficients on arts and religion are barely significant at the 10 percent level.

Column two reports the corresponding coefficients when all subjects are included in the same regression. This specification is less vulnerable to omitted variable bias but might be more vulnerable to measurement error. Sampling error is presumably larger for less common subjects, which suggests that the coefficients on these subjects are more likely to be biased towards zero. Of the seven significant coefficients in column one, four remain significant in column two. Technology increases to 21.7 log points, business and finance drops marginally to 20.5 log points and medicine and history both increase slightly to 11.6 and 10.6 log points, respectively. The only change in conventional thresholds of significance is medicine, which becomes significant at the 10 percent level only. The coefficients on the remaining three subjects that were significant by themselves in column one – social science, arts and religion – fall substantially and become insignificant in column two. Since these are among the most common subjects in the sample, while technology and business and finance are among the least common, it is likely that their significance in column one was driven by omitted variable bias.[25]

Overall, only technology, business and finance, medicine and history are robustly associated with city growth across the two columns. Figure 7 presents binned scatter plots for each of these four subjects, constructed analogously to figure 6 and controlling for all other subjects, i.e. corresponding to the specification from column two of table 5. The larger standard errors translate into greater dispersion around the regression line, especially for technology and business and finance, the least common among the four subjects, but the relationship is clearly visible in all four cases. Technology, business and finance and medicine have a clear interpretation as economically useful knowledge. History, as explained in section II, includes books about current events at the time they were written, which makes it harder to interpret. One possibility is that it captures the existence of a literate elite involved in or at least informed about public life.

It is interesting to further distinguish between books on business and books on finance. The former includes books on management (e.g. accounting, business arithmetic), commerce, transport and communications, while the latter includes books on personal, corporate and public finance. Figure 8 presents separate binned scatter plot for business and finance, controlling for all other subjects, and shows that the relationship with growth is entirely driven by finance books.[26] One interpretation of this finding is that while financial markets, and in particular banking, developed in Europe from the late middle ages onwards, managerial knowledge did not become relevant before the rise of the modern corporation in the second half of the 19th century (e.g. Chandler, 1977).

[25]In addition, two coefficients that were insignificant in column one become negative and significant in column two, agriculture and general. It is plausible that agricultural books were associated with rural growth, which may in turn have been negatively associated with urban growth. The coefficient on general books is harder to interpret, but it should be noted that it is significant at the 10 percent level only. In any case, strong conclusions should not be drawn for these two subjects given the inconsistency between the results in columns one and two.

[26]The data do not have enough power to further distinguish between personal and corporate finance on one hand and public finance on the other: both coefficients are positive and large but insignificant.

Turning to the remaining subjects, books on religion, literature and language, which are presumably associated with the level of literacy in the population, are not associated with growth. Legal books, which might be interpreted as a proxy for the rule of law, are also not significant. Books on social science are significant only when other subjects are omitted. Finally, one finding that seems surprising in light of the economically useful knowledge interpretation is the insignificance of science books, which I turn to next.

#### IV.C.1 The Role of Science

The role of science as a driver of historical innovation and growth is debated. One view (Mathias, 1969; Hall, 1974) is that the Scientific Revolution of the 17th century had little impact on technology in the Industrial Revolution, which was driven by practical innovators devoid of scientific training. In this view, major breakthroughs in physics or chemistry with industrial applications only occurred later, and science only became pivotal in the second stage of industrialization, after 1850. Another view (Musson and Robinson, 1969; Mokyr, 2005*a*) is that the Scientific Revolution led to the diffusion of the scientific mindset associated with the Enlightenment, a belief that material progress can be attained through the accumulation of knowledge. This in turn set off an accumulation of useful knowledge consisting of "catalogs of facts, based on experience and experiment rather than on understanding or careful analysis and testing" (Mokyr, 2005*a*) which drove industrialization, even if the corresponding theoretical breakthroughs only occurred later. The results in the previous section provide some support for both views. Science books are not associated with growth, but books on technology, whose production increased sharply in the 18th century, are. However, the science category analyzed so far combines fields that are typically associated with economic growth – e.g. physics, chemistry or geology – with fields that are less so – e.g. astronomy, botany or zoology, and this aggregation might conceal important heterogeneity. In this section I analyze the relationship between science and growth by field.

I use the LCC subclasses within the Science class Q to divide science into 12 fields: mathematics, astronomy, physics, chemistry, geology, biology, physiology, anatomy, zoology, botany, microbiology and other science. Table 6 shows the correspondence with the DDC and UDC. As for all other subjects above, I estimate two specifications, one where each field enters alone and another where I control for books in all other science fields and subjects. Column one in table 7 reports the results when each field enters alone, and two fields are positive and significant. Chemistry has a coefficient of 18.2 log points, significant at the 5 percent level, and geology has a coefficient of 19.3 log points, significant at the 1 percent level. Both coefficients increase slightly and remain significant in column two, which adds controls for all other science fields and subjects. Chemistry rises to 20.0 log points, significant at the 10 percent level, and geology to 23.1 log points, significant at the 5 percent level. Figure 9 shows the corresponding binned scatter plots. In addition, two fields become negative and significant at the 10 percent level in the second column, anatomy and zoology. In the case of anatomy, including medicine in the regression is likely to be a bad control, in the sense that any effect of anatomy books on growth may work at least partly through the production of medical books. In line with this hypothesis, dropping medicine raises the coefficient on anatomy, which becomes insignificant.[27]

These findings reveal that science did drive growth before and during the Industrial Revolution. While fields like physics and biology may have only developed later, chemistry and geology played an important role in this period. The importance of chemistry in early industrial development is highlighted by Clow and Clow (1952), and that of geology is consistent with the key role of coal mining in the Industrial Revolution.

# V Conclusion

This paper presents evidence on the relationship between knowledge diffusion and economic growth, in line with the Nelson and Phelps (1966) and Schultz (1975) view of human capital. I introduce a new dataset of 5.5 million books published in ten European countries between 1450-1800, and use it to document historical patterns of book production and to investigate the relationship between book production and city growth. I then use information on book subjects to distinguish between interpretations of this relationship. The coefficient is large and robust for books on technology, finance, medicine, history and, within science, chemistry

[27]The same reasoning can be applied to physics and chemistry, on one side, and technology on the other, but dropping technology in the second column only increases the coefficient on physics marginally.

and geology. Other topics, such as religion or literature, are not associated with growth. This suggests that book production reflects the diffusion of economically useful knowledge.

# References

**Acemoglu, Daron, Simon Johnson, and James Robinson.** 2005. "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth." *American economic review*, 546–579.

**Aghion, Philippe, and Peter Howitt.** 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351.

**Allen, R. C.** 2003. "Progress and Poverty in Early Modern Europe." *The Economic History Review*, 56(3): 403–443.

**Bairoch, Paul, Jean Batou, and Chèvre Pierre.** 1988. *Population Des Villes Européennes de 800 à 1850: Banque de Données et Analyse Sommaire Des Résultats (La).* Librairie Droz.

**Barro, Robert J.** 1991. "Economic Growth in a Cross Section of Countries." *The Quarterly Journal of Economics*, 106(2): 407–443.

**Baten, Joerg, and Jan Luiten van Zanden.** 2008. "Book Production and the Onset of Modern Economic Growth." *Journal of Economic Growth*, 13(3): 217–235.

**Benhabib, Jess, and Mark M. Spiegel.** 1994. "The Role of Human Capital in Economic Development Evidence from Aggregate Cross-Country Data." *Journal of Monetary economics*, 34(2): 143–173.

**Blanchard, Olivier Jean, and Lawrence F. Katz.** 1992. "Regional Evolutions." *Brookings papers on economic activity*, 1–75.

**Buringh, Eltjo, and Jan Luiten Van Zanden.** 2009. "Charting the "Rise of the West": Manuscripts and Printed Books in Europe, a Long-Term Perspective from the Sixth through Eighteenth Centuries." *The Journal of Economic History*, 69(02): 409–445.

**Cantoni, Davide, and Noam Yuchtman.** 2014. "Medieval Universities, Legal Institutions, and the Commercial Revolution." *The Quarterly Journal of Economics*, 129(2): 823–887.

**Chandler, A. D.** 1977. *The Visible Hand: The Managerial Revolution in American Business.* Harvard University Press.

**Ciccone, Antonio, and Elias Papaioannou.** 2009. "Human Capital, the Structure of Production, and Growth." *The Review of Economics and Statistics*, 91(1): 66–82.

**Clow, Archibald, and Nan L. Clow.** 1952. *The Chemical Revolution: A Contribution to Social Technology.* Gordon and Breach Science Publishers.

**de Brouillant, Léonce Janmart.** 1888. *Histoire de Pierre du Marteau: imprimeur à Cologne (XVIIe-XVIIIe siècles).* Maison Quantin.

**De Long, J. Bradford, and Andrei Shleifer.** 1993. "Princes and Merchants: European City Growth before the Industrial Revolution." *Journal of Law and Economics*, 671–702.

**Dittmar, J.** 2011. "The Welfare Impact of a New Good: The Printed Book."

**Dittmar, Jeremiah.** 2015. "New Media, Competition and Growth: European Cities after Gutenberg."

**Galor, Oded.** 2005. "From Stagnation to Growth: Unified Growth Theory." *Handbook of economic growth*, 1: 171–293.

**Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer.** 2013. "Human Capital and Regional Development." *The Quarterly Journal of Economics*, 128(1): 105–164.

**Glaeser, Edward L., and Joshua D. Gottlieb.** 2009. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature*, 47(4): 983–1028.

**Glaeser, Edward L., José A. Scheinkman, and Andrei Shleifer.** 1995. "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics*, 36(1): 117–143.

**Grossman, Gene M., and Elhanan Helpman.** 1991. *Innovation and Growth in the Global Economy.* MIT Press.

**Hall, A. Rupert.** 1974. "What Did the Industrial Revolution in Britain Owe to Science?" In *Historical Perspectives: Studies in English Thought and Society, in Honour of J. H. Plumb.* , ed. Neil McKendrick, 129–151. Europa.

**Mathias, Peter.** 1969. "Who Unbound Prometheus? Science and Technical Change, 1600–1800." *Bulletin of Economic Research*, 21(1): 3–16.

**Mitch, David.** 1993. "The Role of Human Capital in the First Industrial Revolution." *The British Industrial Revolution: An Economic Perspective*, 267–307.

**Mokyr, Joel.** 2005*a*. "The Intellectual Origins of Modern Economic Growth." *The Journal of Economic History*, 65(02): 285–351.

**Mokyr, Joel.** 2005*b*. "Long-Term Economic Growth and the History of Technology." In *Handbook of Economic Growth.* Vol. 1, 1113–1180. Elsevier.

**Musson, Albert Edward, and Eric Robinson.** 1969. *Science and Technology in the Industrial Revolution.* Manchester University Press.

**Nelson, Richard R., and Edmund S. Phelps.** 1966. "Investment in Humans, Technological Diffusion, and Economic Growth." *The American Economic Review*, 56(1/2): 69–75.

**Roback, Jennifer.** 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy*, 90(6): 1257–1278.

**Romer, Paul.** 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5): 2.

**Schmitt, C. B., Quentin Skinner, Eckhard Kessler, and Jill Kraye,** ed. 1988. *The Cambridge History of Renaissance Philosophy.* Cambridge:Cambridge University Press.

**Schultz, Theodore W.** 1975. "The Value of the Ability to Deal with Disequilibria." *Journal of economic literature*, 827–846.

**Squicciarini, Mara P., and Nico Voigtländer.** 2014. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." National Bureau of Economic Research.

# VI Appendix

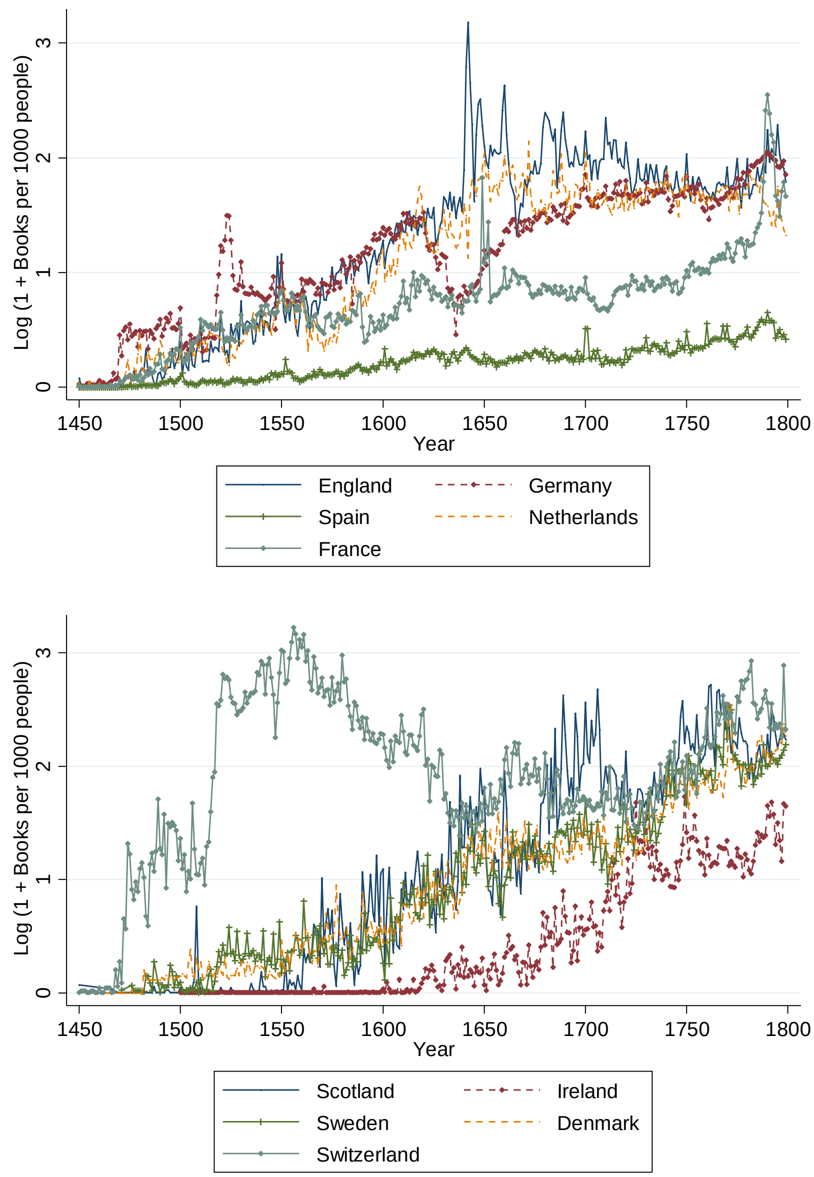
IN PROGRESS

Figure 1: WorldCat Book Record Example

| | |
|---|---|
| **Title:** | **Biblia latina.** |
| **Author(s):** | Gutenberg, Johann,; 1397?-1468. |
| **Publication:** | [Mainz], [Printer of the 42-line Bible (Johann Gutenberg)] |
| **Year:** | 1454-1455? |
| **Description:** | [643] leaves (blank leaves [642] & [643] lacking), bound in 2 volumes 38-40 cm |
| **Language:** | Latin |
| **Contents:** | [1] Leaf [396], containing pts. of chapters III and IV of Jeremiah. Part of columns at outer margin cut off. Rubricated. Bound in ms. leaves.--[2] Leaf [454], containing pts of chap. XLVII and chap. XLVIII of Ezekiel. Part of columns at inner margin and first line of each column cut off.--[3] Leaf [461], containing pts. of chapters VII and VIII of Daniel. Part of columns at inner margin cut off. Gift of Mr. Batchelder. |
| **Standard No:** | **LCCN:** 52-51757 |
| | **SUBJECT(S)** |
| **Genre/Form:** | Incunabula -- Germany -- Mainz -- 1454. |
| **Note(s):** | On paper. Leaves 1-34, 129-158, and 261 are of the first setting. Initial strokes supplied in red on leaves [1]-[130] and on a few other scattered leaves, headlines in alternate red and blue letters, initials in red or blue, some of the larger ones in red and blue, a few with ornaments extending to the margin./ V. 1: 37.4 x 27.3 cm. v. 2: 39.1 x 27.4 cm./ Binding 16-17th century brown polished blind-stamped calf over wooden boards, v. 1 dated 1600, both volumes rebacked in leather approximately of the same color./ Provenance: James Perry, Duke of Sussex, Bishop of Cashel, Earl of Crawford, Earl of Carysfort, Carl H. Pforzheimer./ De Ricci 8. Lazare 40. Norman 41. |
| **Class Descriptors:** | **LC:** BS75 |
| **Other Titles:** | Bible. Latin. Mainz. Gutenberg (42 lines). 1454. |
| **Document Type:** | Book |
| **Entry:** | 19790427 |
| **Update:** | 20140803 |
| **Accession No:** | **OCLC:** 4904654 |

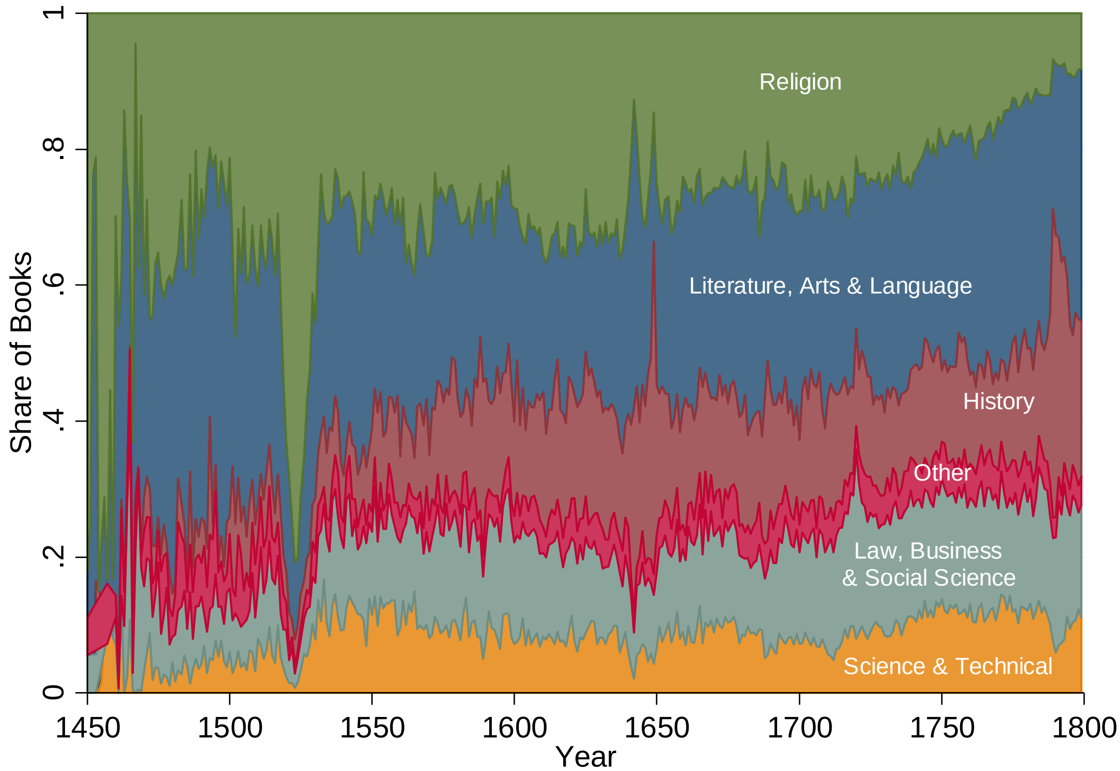
Notes: This figure displays the WorldCat record for the first major book printed in Europe, Gutenberg's Bible. Data on book production are constructed from book records such as this one. The fields used from each record are Publication, which includes the city where the book was printed, Year and Class Descriptors, which includes codes from standard library classification systems that I use to identify the book's subject, when available. In this case the Class Descriptors field includes the Library of Congress Classification BS75, which corresponds to class B "Philosophy, Psychology, Religion" and subclass BS "The Bible". Country of publication is not displayed in book records but is available as a search field, and I obtain it by performing country-specific searches when constructing the dataset.

Figure 2: Per Capita Book Production Across Countries




Notes: This figure presents annual per capita book production for the ten countries in the sample. The numerator includes all book records associated with each country, including the nine percent of records without a city match. The denominator includes urban population only, as measured by the Bairoch data set. When a book record indicates multiple cities of publication, I count one book per city. When the record indicates a range for the year of publication I assign an equal fraction of the book to each year in the range.

Figure 3: Share of Books by Subject




Notes: This figure presents the share of books by subject over time. I combine the 14 subjects defined in table 1 into six broader groups. "Science & Technical" includes Science, Medicine, Technology and Agriculture. "Other" includes General and Philosophy. Book counts for each subject are adjusted by multiplying the raw counts by the ratio of total books to books with subject information at the country-year level, to correct for different levels of subject availability across countries and time.

Figure 4: Per Capita Book Production by Subject (1/2)

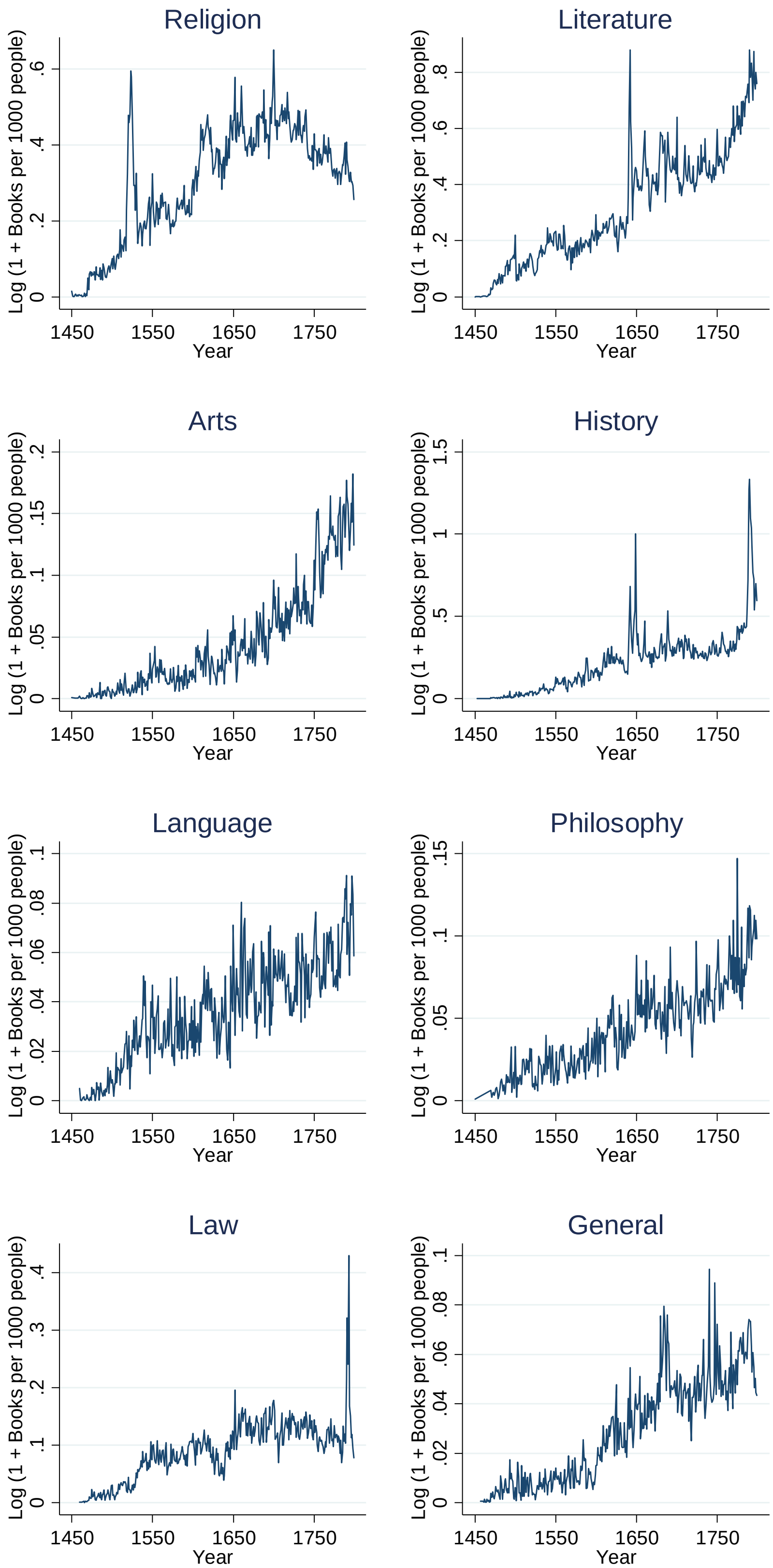
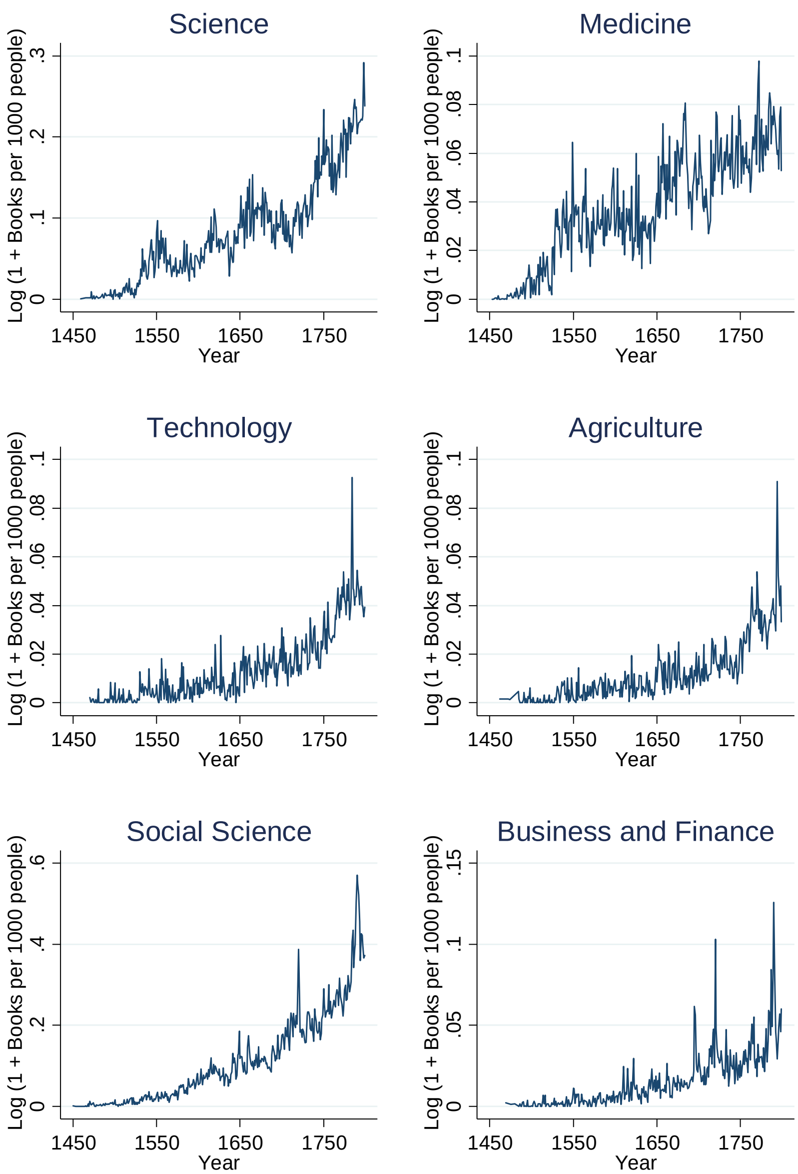
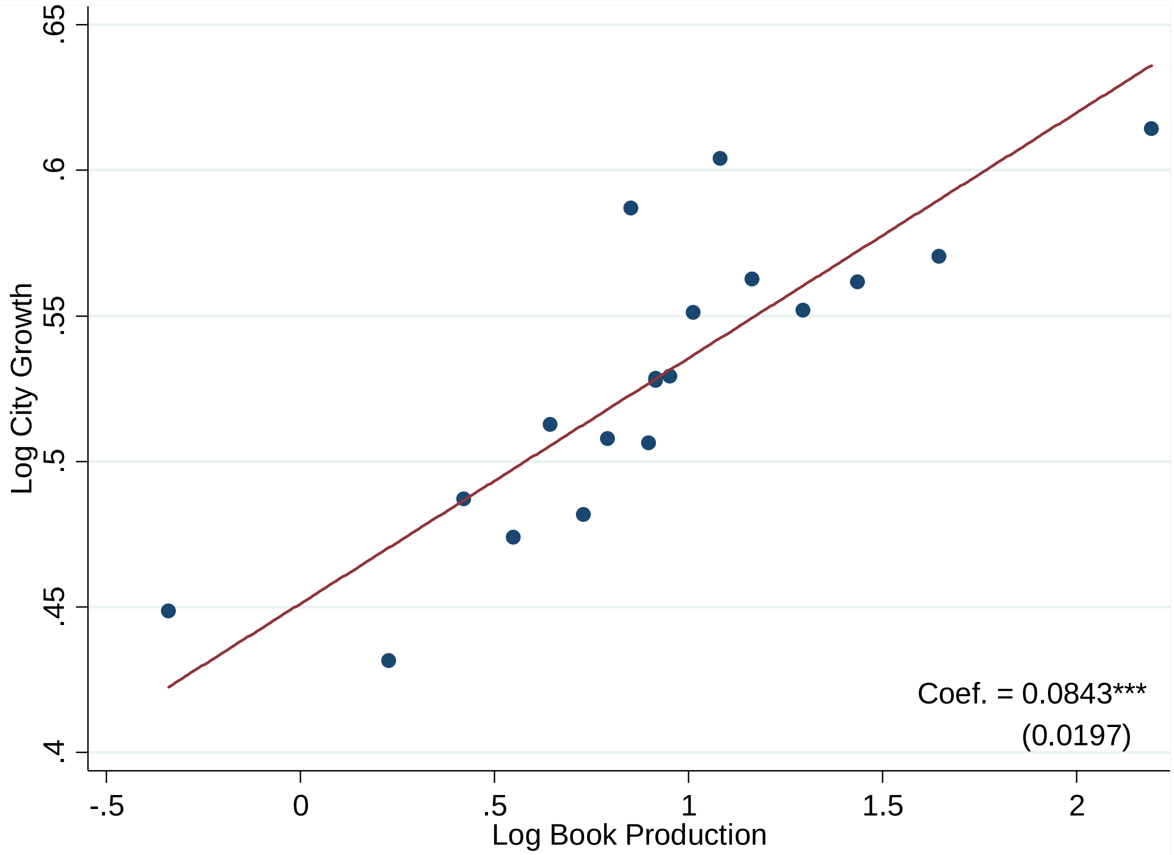


Figure 5: Per Capita Book Production by Subject (2/2)



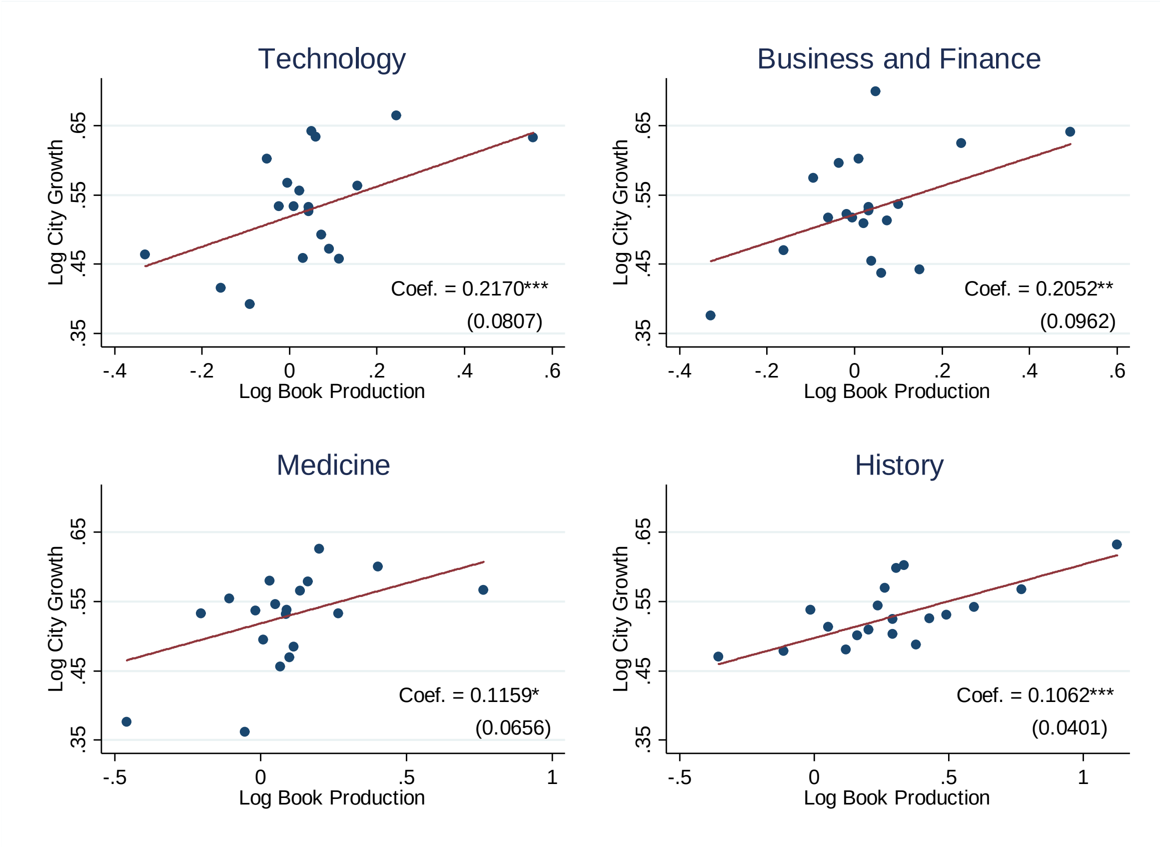

Notes: These figures present annual per capita book production by subject. See notes to figure 2 for details on the measurement of per capita book production. Book counts for each subject are adjusted by multiplying the raw counts by the ratio of total books to books with subject information at the country-year level, to correct for different levels of subject availability across countries and time.

Figure 6: Book Production and City Growth



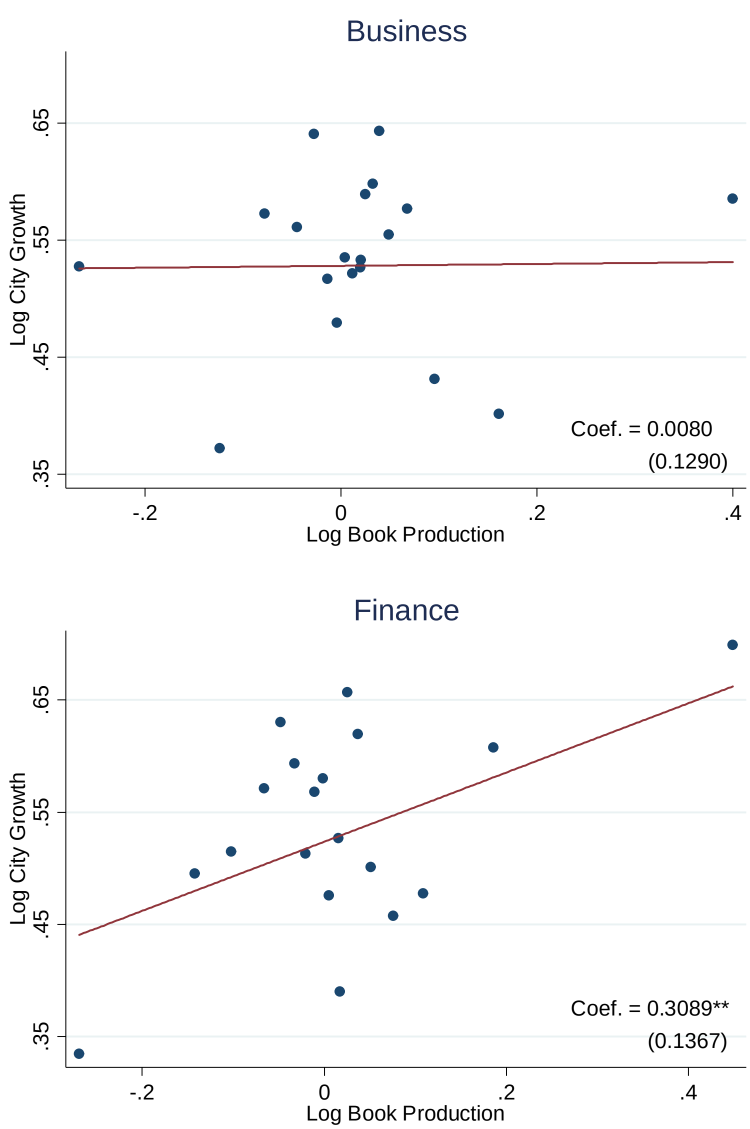

Notes: This figure presents a binned scatter plot of log city growth and log book production. City growth is measured as the change in log population from $t$ to $t + 100$ between 1500 and 1700 and two times the change in log population from $t$ to $t + 50$ between 1700 and 1850. Book production is measured as the stock of unincorporated knowledge at $t$ (see section IV.A for details on the construction of this variable). Both variables are first residualized on city and country-by-year fixed effects and on a quartic in log population at $t$, and then grouped into 19 equal-sized bins for all cities with positive book production at any point during the sample period plus one bin including all cities without book production. The regression line, coefficient and standard error are estimated on the underlying data, not the binned averages. All other binned scatter plots in the paper follow the same procedure.

Figure 7: Book Production and City Growth by Subject



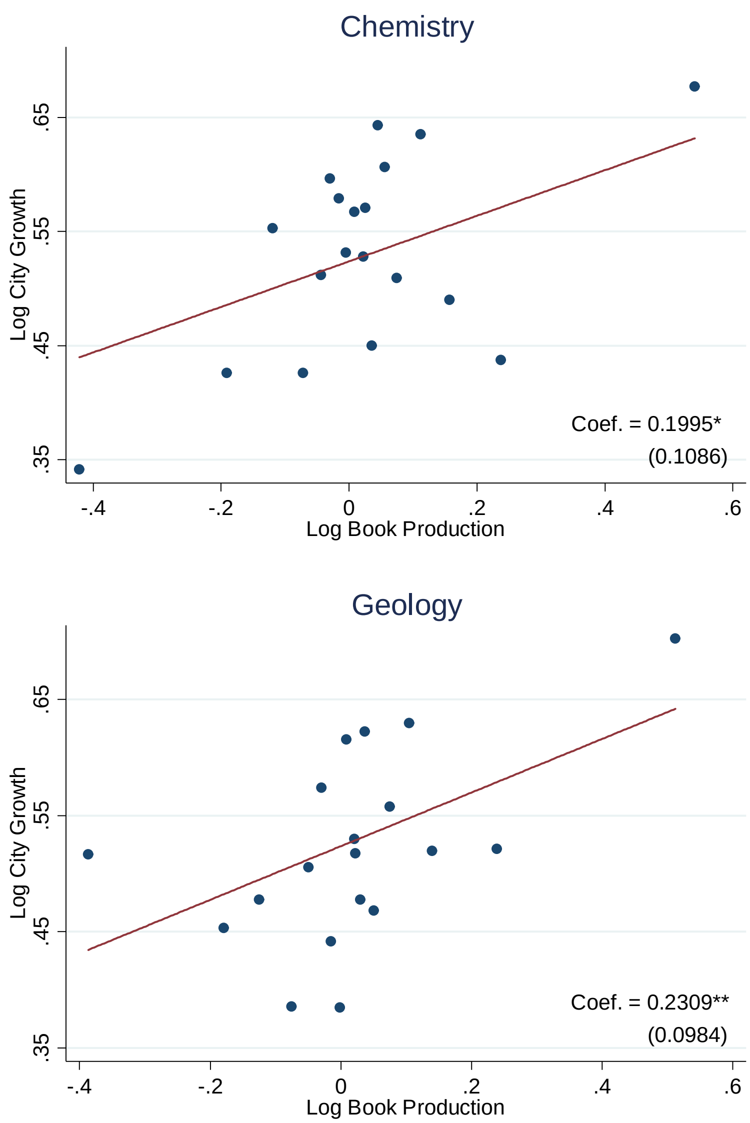

Notes: This figure presents binned scatter plots of log city growth and log book production for books on technology, business and finance, medicine and history. See notes to figure 6 for details on the construction of the variables. The regression line, coefficient and standard error are estimated on the underlying data, not the binned averages.

Figure 8: Book Production and City Growth - Business and Finance




Notes: This figure presents binned scatter plots of log city growth and log book production for books on business and finance. See notes to figure 6 for details on the construction of the variables. The regression line, coefficient and standard error are estimated on the underlying data, not the binned averages.

Figure 9: Book Production and City Growth by Science Field




Notes: This figure presents binned scatter plots of log city growth and log book production for books on chemistry and geology. See notes to figure 6 for details on the construction of the variables. The regression line, coefficient and standard error are estimated on the underlying data, not the binned averages.

Table 1: Classification of Book Subjects

| Subject | Library of Congress | Dewey Decimal Class. | Universal Dec. Class. |
|---|---|---|---|
| Religion | Subclasses BL to BX | Class 2 | Class 2 |
| Literature | Subclasses AC, PA3000 to PA8999, PN to PZ | Class 8 and division 08 | Class 8 (except divisions 80, 81) and division 08 |
| Arts | Classes M, N | Class 7 | Class 7 |
| History | Classes C (except subclass CE), D, E, F and subclass G | Class 9 | Class 9 |
| Language | Subclasses P to PM (except subclass PA3000 to PA8999) | Class 4 | Divisions 80, 81 |
| Philosophy | Subclasses B to BJ | Class 1 except division 15 | Class 1 except division 15 |
| General | Class Z and subclasses AE to AZ | Class 0 except division 08 | Class 0 except division 08 |
| Science | Class Q and subclass CE | Class 5 (except section 526) and sections 611, 612 | Class 5 (except subdivision 528) and subdivisions 611, 612 |
| Medicine | Class R | Division 61 except sections 611, 612 | Division 61 except subdivisions 611, 612 |
| Technology | Class T | Class 6 except divisions 61, 63, 65 | Class 6 except divisions 61, 63, 65 |
| Agriculture | Class S | Division 63 | Division 63 |
| Social Science | Classes G (except subclass G), H (except subclasses HE to HJ), J, L, U, V | Class 3 (except divisions 34, 38 and sections 332, 336), division 15 and section 526 | Class 3 (except division 34 and subdivisions 336, 339), division 15 and subdivision 528 |
| Business and Finance | Subclasses HE to HJ | Divisions 38, 65 and sections 332, 336 | Division 65 and subdivisions 336 and 339 |
| Law | Class K | Division 34 | Division 34 |

Notes: This table presents the correspondence between the 14 book subjects used in the paper and the underlying codes in the Library of Congress Classification (LCC), the Dewey Decimal Classification (DDC) and the Universal Decimal Classification (UDC).

Table 2: Per Capita Book Production by Subject - Correlation Matrix

| | Religion | Literature | Arts | History | Language | Philosophy | Law |
|---|---|---|---|---|---|---|---|
| Religion | 1 | | | | | | |
| Literature | 0.534 | 1 | | | | | |
| Arts | 0.266 | 0.326 | 1 | | | | |
| History | 0.459 | 0.526 | 0.283 | 1 | | | |
| Language | 0.379 | 0.396 | 0.200 | 0.336 | 1 | | |
| Philosophy | 0.413 | 0.431 | 0.219 | 0.362 | 0.323 | 1 | |
| Law | 0.419 | 0.385 | 0.187 | 0.370 | 0.316 | 0.320 | 1 |
| General | 0.304 | 0.340 | 0.183 | 0.305 | 0.270 | 0.271 | 0.247 |
| Science | 0.409 | 0.442 | 0.262 | 0.418 | 0.367 | 0.403 | 0.336 |
| Medicine | 0.348 | 0.383 | 0.194 | 0.352 | 0.325 | 0.339 | 0.374 |
| Technology | 0.191 | 0.224 | 0.175 | 0.220 | 0.182 | 0.195 | 0.156 |
| Agriculture | 0.165 | 0.228 | 0.148 | 0.188 | 0.154 | 0.164 | 0.144 |
| S. Science | 0.407 | 0.471 | 0.243 | 0.454 | 0.315 | 0.381 | 0.370 |
| Bus. Fin. | 0.147 | 0.203 | 0.123 | 0.191 | 0.117 | 0.132 | 0.160 |

| | General | Science | Medicine | Technology | Agriculture | S. Science | Bus. Fin. |
|---|---|---|---|---|---|---|---|
| General | 1 | | | | | | |
| Science | 0.296 | 1 | | | | | |
| Medicine | 0.262 | 0.411 | 1 | | | | |
| Technology | 0.164 | 0.270 | 0.200 | 1 | | | |
| Agriculture | 0.155 | 0.231 | 0.164 | 0.153 | 1 | | |
| S. Science | 0.292 | 0.418 | 0.341 | 0.254 | 0.217 | 1 | |
| Bus. Fin. | 0.135 | 0.159 | 0.140 | 0.0963 | 0.101 | 0.220 | 1 |

Notes: This table presents the correlation matrix per capita book production across subjects. Observations are at the city-year level.

Table 3: Book Production and City Growth

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | | | | | Books> 0 | Weighted | |
| Log Books, $\gamma = 0.1$ | 0.0843*** | | | 0.0556* | 0.0864*** | 0.1112*** | |
| | (0.0197) | | | (0.0300) | (0.0211) | (0.0292) | |
| Log Books, $\gamma = 0.2$ | | 0.0965*** | | | | | |
| | | (0.0220) | | | | | |
| Log Books, $\gamma = 0.05$ | | | 0.0714*** | | | | |
| | | | (0.0184) | | | | |
| Books, $\gamma = 0.1$ | | | | | | | 0.0008** |
| | | | | | | | (0.0003) |
| Log Initial Population | Y | Y | Y | Y | Y | Y | Y |
| Country x Year FE | Y | Y | Y | Y | Y | Y | Y |
| City FE | Y | Y | Y | Y | Y | Y | Y |
| City Time Trends | | | | Y | | | |
| Observations | 3060 | 3060 | 3060 | 3060 | 2217 | 3060 | 3060 |
| Number of Cities | 1119 | 1119 | 1119 | 1119 | 877 | 1119 | 1119 |

Notes: This table presents regressions of city population growth on book production. City growth is measured as the change in log population from t to t+100 between 1500 and 1700 and two times the change in log population from t to t+50 between 1700 and 1850. Book production is measured as the stock of unincorporated knowledge at t (see section IV.A for details on the construction of this variable). Columns one to three use alternative assumptions about the rate of knowledge incorporation into output. Column four adds city-level time trends. Column five restricts the sample to observations with positive book production, and column six weights cities by their average population in the sample period. Column seven uses the level rather than the log of book production.

Table 4: Book Production and City Growth by Time Period and Country

| | (1) All | (2) France | (3) Germany | (4) England | (5) Netherlands | (6) Spain | (7) Other |
|---|---|---|---|---|---|---|---|
| Log Books | | 0.1564*** | 0.0922*** | -0.0102 | 0.1307** | 0.1351* | 0.0736** |
| | | (0.0446) | (0.0322) | (0.0883) | (0.0606) | (0.0799) | (0.0359) |
| Log Books x 1500 | 0.1085*** | | | | | | |
| | (0.0360) | | | | | | |
| Log Books x 1600 | 0.1244*** | | | | | | |
| | (0.0243) | | | | | | |
| Log Books x 1700 | 0.1096*** | | | | | | |
| | (0.0302) | | | | | | |
| Log Books x 1750 | 0.0113 | | | | | | |
| | (0.0262) | | | | | | |
| Log Books x 1800 | 0.0895*** | | | | | | |
| | (0.0257) | | | | | | |
| Log Initial Population | Y | Y | Y | Y | Y | Y | Y |
| Country x Year FE | Y | Y | Y | Y | Y | Y | Y |
| City FE | Y | Y | Y | Y | Y | Y | Y |
| Observations | 3060 | 790 | 700 | 364 | 171 | 769 | 266 |
| Number of Cities | 1119 | 339 | 237 | 156 | 60 | 252 | 75 |

Notes: This table presents additional regressions of city population growth on book production. See notes to table 3 for details on the construction of the variables. Column one interacts book production with time indicators. Columns two through seven split the sample by geography.

Table 5: Book Production and City Growth by Subject

| | (1) Univariate | (2) Multivariate |
|---|---|---|
| Religion | 0.042* | 0.006 |
| | (0.025) | (0.035) |
| Literature | 0.036 | -0.007 |
| | (0.024) | (0.035) |
| Arts | 0.086* | 0.053 |
| | (0.051) | (0.063) |
| History | 0.093*** | 0.106*** |
| | (0.029) | (0.040) |
| Language | 0.052 | -0.019 |
| | (0.043) | (0.062) |
| Philosophy | 0.035 | -0.054 |
| | (0.044) | (0.069) |
| Law | 0.031 | -0.034 |
| | (0.032) | (0.037) |
| General | 0.007 | -0.111* |
| | (0.049) | (0.064) |
| Science | 0.039 | -0.069 |
| | (0.031) | (0.050) |
| Medicine | 0.100** | 0.116* |
| | (0.041) | (0.066) |
| Technology | 0.172*** | 0.217*** |
| | (0.056) | (0.081) |
| Agriculture | -0.006 | -0.215** |
| | (0.071) | (0.093) |
| Social Science | 0.072** | 0.015 |
| | (0.035) | (0.057) |
| Business and Finance | 0.215** | 0.205** |
| | (0.087) | (0.096) |
| Log Initial Population | Y | Y |
| Country x Year FE | Y | Y |
| City FE | Y | Y |
| Observations | 3060 | 3060 |
| Number of Cities | 1119 | 1119 |

Notes: This table presents regressions of city population growth on book production by subject. See notes to table 3 for details on the construction of the variables. In column one each coefficient is estimated in a separate regression where the corresponding subject enters alone, while column two reports coefficients when all subjects are included in the same regression.

Table 6: Classification of Science Books

| Field | Library of Congress | Dewey Decimal Class. | Universal Dec. Class. |
|---|---|---|---|
| Mathematics | Subclass QA | Division 51 | Division 51 |
| Astronomy | Subclasses QB, CE | Division 52 except section 526 | Division 52 except subdivision 528 |
| Physics | Subclass QC | Division 53 | Division 53 |
| Chemistry | Subclass QD | Division 54 | Division 54 |
| Geology | Subclass QE | Divisions 55, 56 | Divisions 55, 56 |
| Biology | Subclass QH | Division 57 except sections 571, 572, 573, 575, 579 | Division 57 except subdivisions 577, 578, 579 |
| Physiology | Subclass QP | Sections 571, 572, 573, 575, 612 | Subdivisions 577, 612 |
| Anatomy | Subclass QM | Section 611 | Subdivision 611 |
| Zoology | Subclass QL | Division 59 | Division 59 |
| Botany | Subclass QK | Division 58 | Division 58 |
| Microbiology | Subclass QR | Section 579 | Subdivisions 578, 579 |
| Other Science | Subclass Q | Division 50 | Division 50 |

Notes: This table presents the correspondence between the 12 scientific fields used in the paper and the underlying codes in the Library of Congress Classification (LCC), the Dewey Decimal Classification (DDC) and the Universal Decimal Classification (UDC).

Table 7: Book Production and City Growth by Science Field

| | (1) Univariate | (2) Multivariate |
|---|---|---|
| Mathematics | 0.005 | -0.059 |
| | (0.063) | (0.077) |
| Astronomy | 0.089 | 0.003 |
| | (0.061) | (0.079) |
| Physics | -0.008 | -0.127 |
| | (0.062) | (0.088) |
| Chemistry | 0.182** | 0.200* |
| | (0.078) | (0.109) |
| Geology | 0.193*** | 0.231** |
| | (0.062) | (0.098) |
| Biology | -0.037 | -0.071 |
| | (0.053) | (0.098) |
| Physiology | -0.054 | -0.115 |
| | (0.092) | (0.106) |
| Anatomy | -0.053 | -0.189* |
| | (0.075) | (0.102) |
| Zoology | 0.011 | -0.120* |
| | (0.057) | (0.070) |
| Botany | 0.086 | 0.060 |
| | (0.055) | (0.087) |
| Microbiology | 0.207 | 0.332 |
| | (1.408) | (1.311) |
| Other Science | 0.069 | -0.010 |
| | (0.060) | (0.096) |
| Log Initial Population | Y | Y |
| Country x Year FE | Y | Y |
| City FE | Y | Y |
| Observations | 3060 | 3060 |
| Number of Cities | 1119 | 1119 |

Notes: This table presents regressions of city population growth on book production by scientific field. See notes to table 3 for details on the construction of the variables. In column one each coefficient is estimated in a separate regression where the corresponding field enters alone, while column two reports coefficients when all scientific fields, as well all other subjects apart from science, are included in one regression.